

Computer Science

REPRESENTING MULTIMEDIA RESOURCE CONTENT ON THE SEMANTIC WEB: FRAMENET, VEML, AND THE SEMANTIC WEB

Matthew S. Gerber (m-gerber@onu.edu), Jerome A. Feldman (jfeldman@icsi.berkeley.edu)*

University of California, Berkeley, California 94720

Many projects in the field of computer science interact directly with large databases of information pertinent to the task at hand. Indeed, some research efforts have as their only goal the development of these resources with the expectation that they will be used by other groups in a related field. A major concern of such resource development projects is how the data will be distributed to other groups for further use. The semantic web research effort at the World Wide Web Consortium has developed a number of representation languages for use in sharing, reusing, and describing online resources, the most recent of which is the web ontology language OWL. OWL provides a rich framework for resource description and interoperability, and abstracted below is its application to data from two resource development groups – FrameNet, at the University of California, Berkeley and the ARDA Challenge Project on Event Taxonomy.

The FrameNet group is in the midst of expanding and refining a database that provides, for a substantial portion of the English language, semantic specifications of words and the roles they play in sentences. Areas of potential use for FrameNet data include natural language processing, lexicographical reference and research, and language teaching. In the past, this data was distributed in a number of XML files that were structured in a disjoint, cumbersome way. A mapping from XML to the OWL representation language has been created. The XML-OWL mapping consists of the definition of a class hierarchy for basic elements of the database as well as the definition and attachment of descriptive properties that convey information contained in the database. An OWL version of FrameNet data allows any amount of the data to be extracted, manipulated, and integrated with other OWL-encoded data by using efficient, readily available tools designed for semantic web languages.

The Project on Event Taxonomy is aimed at developing a resource similar to the FrameNet database but in the domain of video surveillance data. Tools have been created that allow for the manual (human-based) annotation of events and objects present in video streams. Video Event Markup Language (VEML) is a formal representation language designed for organizing and sharing these annotations, but like the FrameNet database it is structured in an arbitrary manner that requires custom programming from those who use the data. The VEML-OWL mapping gives VEML depictions of videos the benefits of a widely standardized format.

The two projects mentioned above were selected because they come from separate domains and thus show the utility of a standardized representation language such as OWL. OWL-encoded data from the two projects above can be seamlessly integrated and made available to other applications and research groups in a convenient format. Wider and more efficient use of the data is possible and expected.